DOCUMENT RESUME

ED 400 285 TM 025 553

AUTHOR Gershon, Richard C.; Bergstrom, Betty
TITLE Individual Differences in Computer Adaptive Testing: Anxiety, Computer Literacy and Satisfaction.
PUB DATE Apr 91
NOTE 18p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 21-23, 1991).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adaptive Testing; *Adults; *Computer Assisted Testing; Computer Literacy; *Individual Differences; *Satisfaction; *Test Anxiety; Test Length

ABSTRACT

The relationship of several individual differences variables to Computer Adaptive Testing (CAT) as compared with traditional written tests are explored. Seven hundred sixty-five examinees took a Computer Adaptive Test and two fixed-length written tests. Each examinee also answered a computer literacy inventory, a satisfaction questionnaire, and a test anxiety survey. Test anxiety was found to be a significant factor in performance on both of the written tests, but not on the CAT test. Anxiety was also found to be a significant factor on several of the items on the satisfaction questionnaire. Overall, significant factors that predict satisfaction with CAT testing included level of test anxiety, computer literacy, and test length (the CAT test varied in terms of the number of items administered). Results are discussed in terms of the political and practical implications of administering CAT tests as compared to administering traditional written tests. The results also indicate that some of the individual differences variables that have been found to affect performance on written tests are not significant in CAT. (Contains two tables and six references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 400 285

Individual Differences in Computer Adaptive Testing:

Anxiety, Computer Literacy and Satisfaction

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL
HAS BEEN GRANTED BY
BETTY BERGSTROM

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

Richard C. Gershon

Northwestern University
and
Computer Adaptive Technologies

Betty Bergstrom

American Society of Clinical Pathologists

Paper presented at the Annual Meeting of
the National Council of Measurement in Education
April, 1991

BEST COPY AVAILABLE

TM025553

ABSTRACT

The relationship of several individual differences variables with computer adaptive testing (CAT) versus traditional written tests are explored. Seven-hundred sixty-five examinees took a computer adaptive test and two fixed length written tests. Each examinee also answered a computer literacy inventory, a satisfaction questionnaire, and a test anxiety survey. Test anxiety was found to be a significant factor in performance on both of the written tests, but not on the CAT test. Anxiety was also found to be a significant factor on several of the items on the satisfaction questionnaire. Overall, significant factors which predict satisfaction with CAT testing included level of test anxiety, computer literacy, and test length (the CAT test varied in terms of the number of items administered). Results are discussed in terms of the political and practical implications of administering CAT tests as compared to administering traditional written tests. The results also indicate that some of the individual differences variables which have been found to affect performance on written tests are not-significant in CAT.

In recent years, researchers have noted a wide array of improvements which can be gained by switching from classical test theory to Item Response Theory (IRT). Computer adaptive testing (CAT) utilizes the concepts of IRT to decrease test length, improve reliability, control test difficulty, and in general, to bring testing into the twenty-first century. Unfortunately, some of these advances have been made without regard to the lessons learned by proponents of classical test theory. These issues relate not only to classical psychometric issues, but also to individual differences. For example, it is widely accepted that there is a relationship between test anxiety and performance on achievement based tests (e.g. Deffenbacher, 1978). But how will test anxiety impact on IRT based CAT tests? There are also other individual-differences variables which should be considered in CAT. Does one's familiarity with computers improve one's performance when a test is given on a computer? How do all of the individual differences variables affect satisfaction with CAT versus written test administrations?

While test anxiety has been shown to be a significant component in predicting performance on traditional tests, only recently has this effect been demonstrated with IRT based written tests (Gershon, 1991). The IRT based study of written tests demonstrated that test anxiety interacts with test performance in two different ways. First, there is a main effect for anxiety. Low-anxious persons perform better than high anxious persons overall. Second, within test performance varies depending on one's level of test anxiety. When high-anxious persons are

presented with difficult items at the beginning of the test they perform better than expected, but that performance falls off towards the end of the test. The opposite is true for low-anxious persons who improve as test length increases. These findings may have foreboding consequences for CAT. For instance, if test length is significantly shortened, then low anxious persons may not have a sufficient warm-up period to reach their maximum performance level. This effect may be further compounded in CAT tests which usually control test difficulty so that even the highest-able persons answer only 50% of the items correctly (most CAT algorithms target test items to the ability of the examinee, effectively controlling the percentage of items which a person answers correctly). This may be discouraging and result in decreased performance measures. Clearly the effect may be further impact upon an examinee who is also extremely test anxious.

The exponential consequence of knowledge and familiarity with computers on actual test performance must also be considered. While there is little worry as to the impact of familiarity with using a number-two pencil on performance on a written test, can the same be said for performance on a test administered by computer? Koslowksy, Lazar and Hoffman (1988) have explored the concept of computer literacy, but only in relationship to the likelihood of actually using computers.

At this stage in the history of CAT, the field faces a myriad of consequences not raised in written testing situations. Unanswered issues in this regard include, but are not limited to: What variables effect satisfaction with CAT testing? Can

the testing situation be modified in ways which improve satisfaction without jeopardizing the integrity of the tests? Are there situations in which CAT tests are preferred by examinees over the use of written tests? And, what factors predict apprehension before taking a computerized test?

Method

CAT test. A total of 765 examinees were administered a variable-length computer adaptive test and two fixed length written tests. The CAT test was variable in length and drew from a pool of 726 items. The test specifications included a confidence-interval based stopping rule relative to making a pass-fail decision. The examination continued until a person's ability estimate was clearly above or below the pass-fail point, resulting in a variable length test. People whose ability measures are near the pass-fail point take longer tests than persons whose ability places them in the clear pass or clear fail range of ability (see Bergstrom and Lunz, 1991).

Random assignment was also made for the probability of making a correct response. The CAT algorithm keeps track of the current predicted ability level of the examinee in order to select the next item. While traditional CAT tests have offered items at the 50% likelihood of a correct response, some subjects were assigned to 60% and 70% conditions for this study (see Gershon and Bergstrom, 1990).

Persons were also randomly assigned to receive a very easy item first, a very hard item first, or an item of average difficulty as the first item administered in the adaptive test. Traditionally the initial item in adaptive tests has been given at either the mean population ability or at the pass/fail point.

Another factor which was not explicitly randomly assigned, but proved to be a significant factor in many of the analyses which follows is the total number of items administered to an individual. The total number of items administered over two CAT tests taken in succession ranged from 50 to 480. Persons were randomly assigned to minimum test length conditions of either 50 or 100 items. Persons were also, unbeknownst to them, administered a retest if they completed their first test with time to spare in the four hour maximum testing time. Therefore, it was possible for a person who took a maximum length test to be immediately readministered another 240 item test. At the other extreme, some persons were never given a retest and therefore could have completed their test in 50 items. In summary, the total number of items administered refers to the total number of unique CAT items shown to the individual in the one sitting.

Surveys. Each person also answered a eight-item computer literacy inventory (based on Koslowsky, Lazar & Hoffman, 1989) and most of the examinees answered a 12-item test anxiety survey based on Mandler & Sarason's Test Anxiety Scale (1952). The two scales were randomly assigned to be administered by the computer before or after the administration of the actual CAT test questions.

The computer literacy items asked questions such as "Whenever I use a computer, I am afraid I will break it" and "I prefer using a word processor to a typewriter."

The test anxiety survey measured a person's level of trait test anxiety. This would be the level of anxiety one is subject to in all types of testing situations, regardless of the content area. This true-false survey asked questions such as "I have an uneasy upset feeling before taking an important test" and "Thoughts of doing poorly interfere with my performance on tests."

Written Tests. The short written test consisted of 109 items drawn from the same item pool as the CAT test and was administered before the CAT test for some examinees and after the CAT test for others, either on the same day or up to several weeks apart. A six-item satisfaction questionnaire was administered following the second of these two tests.

The long written test was always administered last and usually several weeks after the completion of the first two tests. This second written test consisted of 189 items with comparable test specifications to the other two test, but no overlapping items. It is believed that the examinees treated the last test the most seriously, and therefore the results from the long test are probably the best predictors of a person's true ability.

Results

Factors of Ability

Anxiety. An initial analysis of variance was performed to determine whether the timing of the administration of the test anxiety questionnaire was a factor in the test anxiety score. The results were not significant indicating, that there was no difference if the questionnaire was administered before or after the adaptive test.

Using analysis of variance, anxiety was found to be a significant variable in predicting performance on all three tests (for the CAT test $R^2 = .05$, $F_{(1,583)} = 31.66$, $p < .001$); for the short written test $R^2 = .05$, $F_{(1,592)} = 29.45$, $p < .001$); and for the long written test $R^2 = .05$, $F_{(1,552)} = 39.2$, $p < .001$). These findings indicate that anxiety impacts performance similarly in all three formats. However, two additional factors need to be considered. First, the anxiety questionnaire was given in conjunction with the CAT test, and therefore, the anxiety score should have the strongest relationship with the CAT score (however, it does not). Second, since previous research has shown that anxiety and ability are so closely related, one must further examine the analyses listed above without including the ability component of anxiety. In this regard, two additional analyses were completed. In the first, an analysis of covariance was performed with the dependent variable being the score obtained on the CAT test, the independent variable of anxiety, and the covariate of ability as tested on the long test. In this case the impact of anxiety was *not* significant ($R^2 = .36$;

Anxiety: $F_{(2,700)} = 1.61$, N.S.; Ability: $F_{(1,700)} = 363.93$, $p < .001$). In other words, when looking only at the unique aspect of anxiety which is not related to ability, anxiety is not a component in measured ability on a CAT test. This was not the case when the unique anxiety component is analyzed in relationship to the score obtained on the short written test ($R^2 = .42$; Anxiety: $F_{(2,709)} = 4.65$, $p < .01$; Ability: $F_{(1,709)} = 473.34$, $p < .001$).

This is an extremely important finding in that most previous studies of test anxiety have assumed that test anxiety was inextricably related to performance. Traditionally, high-able individuals were generally considered less test anxious because they were likely to exhibit superior performance and thus they had less need to be anxious. Similarly, high anxious people were thought to be justified in their beliefs due to their low ability. This may prove to be the first study which has demonstrated than an ability test can be administered which is not subject to the differential effects of test anxiety.

While it is a relatively simple task to observe the impact of anxiety on total test measures, it is also important to begin to understand the impact of anxiety on within test performance. What is happening in the CAT test that is not impacted by the effect of anxiety? Since we already know that persons on written tests vary in their within test performance based upon their level of test anxiety, one would assume that the same is true of persons taking an adaptive test. To look at this hypothesis, Rasch ability measures were calculated for each group of ten successive items within the CAT test. Analysis of variance was used to analyze

the impact of anxiety on each of these sub-test measures. Anxiety was a significant factor in the ability measure obtained from the first ten items only ($R^2 = .02$, $F_{(1,490)} = 11.94$, $p < .001$), the analyses of all subsequent groups of items was not significant. Therefore, the adaptive testing algorithm, while initially subject to the same anxiety component found in most written tests, is able to not only to zero in on a persons ability measure, but do so also while seemingly eliminating the influence of test anxiety.

Test Difficulty. It should be noted that several additional variables were not found to have a significant impact on person ability measures. For instance, the overall difficulty of the test was not found to be a significant factor in predicting person ability measure using analysis of variance. There were no systematic differences in ability measure based upon whether the person taking the test was administered items at the 50% probability of a correct response, the 60% probability of a correct response, or the 70% probability of a correct response. Another analysis of variance was conducted to determine whether the difficulty of the test interacted with a persons level of anxiety. Again the results were not significant.

Initial test item. Participants in the study were also randomly assigned to receive a very easy item first, a very hard item first, or an item of average difficulty as the first item administered in the adaptive test. An analysis of variance equation failed to obtain significance for the starting difficulty of the test in

predicting the final test score. There was also no significant interaction of the initial item difficulty with anxiety on the total test score.

Satisfaction Variables

The answers to the questions in the satisfaction questionnaire were each examined using analysis of variance equations which examined the effects of test anxiety, computer literacy, the total number of test items administered, the starting difficulty of the test, the overall difficulty of the test (50%, 60%, or 70% probability of a correct response), and the person's ability measure. Interaction effects were also checked for anxiety with starting difficulty, anxiety with overall difficulty, and overall difficulty with starting difficulty. Most of the interaction effects did not add significantly to the variance accounted for, and are therefore not listed. All of the items were answered on a likert type scale where 5 = strongly agree, 4 = agree, 3 = neutral, 2 = disagree, and 1 = strongly disagree. Table 1 shows the correlation matrix for the continuous variables used to predict satisfaction. Table 2 shows the correlation matrix for the continuous variables used to predict satisfaction.

Question 1. The first satisfaction question looked at a persons preference for taking a computerized test over taking a paper and pencil based test. Only the literacy and total number of items variables were found to be significant using analysis of variance (R^2 = .06, $F_{(2,703)}$ = 20.53, p < .001). Therefore, persons

Table 1

Significant Indpendent Variables Using Analysis of Variance or Covariance

	Ability Measure	Anxiety	Computer Literacy	Total Number Items	Test Diff.	Anxiety X Test Diff.
1) I liked CAT better			X	X		
2) I like paper and pencil better		X	X	X		
3) I like both equally			X			
4) I was apprehensive at the beginning of CAT		X	X			
5) I was apprehensive at the end of CAT	X	X		X	X	X
6) I would like [future certification] exams to be on the computer			X	X		

Table 2

Mean Response of Satisfaction Questions and Correlations of Continuous Variables with Satisfaction

	Mean Response	Ability Measure	Anxiety	Computer Literacy	Total Number of Items
1) I liked CAT better	3.00	.11$^{=}$	-.12$^{=}$	.19$^{=}$	-.15$^{=}$
2) I like paper and pencil better	3.13	-.13$^{=}$	.12**	-.15$^{=}$	.14$^{=}$
3) I like both equally well	2.53	.05	-.03	.15$^{=}$	-.08*
4) I was apprehensive at the beginning of CAT	2.89	-.12$^{=}$	.45$^{=}$	-.21$^{=}$	-.05
5) I was apprehensive at the end of CAT	2.61	-.31$^{=}$	.35$^{=}$	-.16$^{=}$	-.14$^{=}$
6) I would like [future certification] exams to be on the computer	2.69	.10$^{=}$	.12**	-.23$^{=}$	-.10$^{=}$

NOTE: $^{=}$ $p < .001$; $^{-}$ $p < .01$; * $p < .05$

BEST COPY AVAILABLE

who are more computer literate, and those who took fewer items, were more likely to endorse taking the computer based test over the paper and pencil test. It should be noted, however, that only 5% of the variance in the satisfaction question was determined by these two variables.

Question 2. The second question appeared to be the opposite of the first, but several additional factors were found to be meaningful. Computer literacy, total items administered, test anxiety, and type of test were found to be significant factors using analysis of variance ($R^2 = .05$, $F_{(3,518)} = 8.28$, $p < .001$)). In general less computer literate persons preferred the paper and pencil test. This was also the case for those who are more test anxious and those persons who took longer computer tests.

Question 3. The explored variables could only account for two percent of the variance in persons who rated the question "I like both tests equally well" ($F_{(1,703)} = 15.12$, $p < .001$). Indeed the only factor that could help predict this question was a person's degree of computer literacy; persons higher in computer literacy were more likely to endorse liking both test formats equally well.

Question 4. The results of Question Four greatly help to clarify the factors involved in the apprehension which persons have prior to taking a CAT test. The influences of anxiety and computer literacy accounted for a total of 22% of the variance in the responses ($F_{(3,504)} = 54.11$, $p < .001$). Highly test anxious persons were the most likely to be apprehensive, while some degree of computer literacy corresponded to less apprehension. The significant correlation with ability

(see Table 1) would also indicate that more able persons would be less apprehensive going into a CAT test, however, based on the results of the analysis of variance it appears that the component in ability which is related to apprehension is better conceptualized in terms of the decreased level of test anxiety one tends to find in high able persons.

Question 5. This question addressed apprehension at the end of the CAT test. A full 23% of the variance in responding to this question was accounted for, but this time using a new group of factors as compared to pre-test apprehension. Test anxiety was also the greatest factor in predicting apprehension following the CAT test. The next major factor was ability as measured on the CAT test. Subjects who performed well on the test were less likely to be apprehensive following the test than were their less-able counterparts.

Question 5

Analysis of Variance

Source	SS	DF	MS	F	P
Anxiety	62.67	1	62.87	62.85	0.001
Difficulty	12.31	2	6.16	6.23	0.01
# of Items	7.16	1	7.16	7.25	0.01
Ability	37.03	1	37.03	37.48	0.001
Anxiety X Difficulty	8.35	1	4.17	4.22	0.05
Error	513.84	520	1.00		

Apprehension following the test was further moderated by the probability of a correct response for the given test. Interestingly enough, apriori contrasts showed that persons who had a 70% probability of a correct response were more

apprehensive following the test than were those who received items at the 50% probability. There was no significant difference between apprehension of those in the 60% probability group as compared to the 50% group, or as compared to the 70% group.

The next greatest factor which helped to predict this question was a interaction between level of test anxiety and the probability of a correct response. In general, persons in the 50% probability group were less apprehensive than those in the easier test conditions at almost every anxiety level. Also the difference in level of anxiety between the least and most test anxious individuals was minimized for persons in the 50% group. Finally, apprehension following the test was also effected by the total number of items administered. The more items a person received, the more apprehensive they were following the test.

Question 6. The final satisfaction question asked if the individual would prefer to take their actual certification examination on the computer. Literacy and the total number of items administered predicted six percent of the variance in responding to this question ($F_{(2,694)} = 23.03$, $p < .001$). Persons who were more computer literate were more likely to endorse the idea of the computer administered test, as were those who took fewer items.

Discussion

Individual differences variables play roles in computer adaptive testing which are not necessarily predicted by previous studies that utilized classical or IRT based

written tests. In general, CAT has been shown to help eliminate the effects of what have traditionally been considered to be confounding variables. Even so, anxiety in particular continues to play a role in the testing procedure relative to perceived satisfaction. New confounding variables including test length and computer literacy should also be considered in future research studies surrounding the use of IRT and CAT.

REFERENCES

Bergstrom, B., Lunz, M., and Gershon, R. (1991). Altering the level of difficulty in computer adaptive testing. Paper presented at the annual meeting of the National Council of Measurement in Education in Chicago.

Deffenbacher, J. L. (1978). Worry, emotionality, and task-generated interference in test anxiety: an empirical test of attentional theory. *Journal of Educational Psychology*, 70, 248-254.

Gershon, R. C. (1991). Test anxiety and item order: new concerns for item response theory. In M. Wilson (Ed.), *Objective Measurement: Theory to Practice*. Norwood, NJ: Ablex.

Koslowsky, M., Lazar, A., and Hoffman, M. (1988). Validating an attitude toward computer scale. *Educational and Psychological Measurement*, 48, 517-521.

Lunz, M. and Bergstrom, B. (1991). Confidence in pass-fail decisions for computer adaptive and paper and pencil examinations. Paper presented at the annual meeting of the American Educational Research Association in Chicago.

Mandler, G. & Sarason, S.B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology*, 47, 166-173.